Forward Reliability Markets:
Less Risk, Less Market Power, More Efficiency

Peter Cramton and Steven Stoft
September 2007

Abstract

A forward reliability market is presented. The market coordinates new entry through the forward procurement of reliability options—physical capacity bundled with a financial option to supply energy above a strike price. The market assures adequate generating resources and prices capacity from the bids of competitive new entry in an annual auction. Efficient performance incentives are maintained from a load-following obligation to supply energy above the strike price. The capacity payment fully hedges load from high spot prices, and reduces supplier risk as well. Market power is reduced in the spot market, since suppliers enter the spot market with a nearly balanced position in times of scarcity. Market power in the reliability market is addressed by not allowing existing supply to impact the capacity price. The approach, which has been adopted in New England and Colombia, is readily adapted to either a thermal or a hydro system.

1 Understanding the generation adequacy problem

Generation adequacy is one component of reliability, and the only component referred to whenever "reliability" is mentioned in this paper. Obtaining adequate capacity is the subject of the paper, but the importance of the prescribed reliability options approach is that it nearly solves the three worst problems of contemporary electricity markets: investment risk, market power and inefficient pricing. Contrary to conventional wisdom, the level of generation adequacy is not much of a problem because marginal generating capacity is relatively inexpensive when compared with other costs of delivered energy, and because, as with most optima, the derivative of cost with respect to capacity is zero at the optimal capacity level.

The true cost of the adequacy problem has been the distortion of market designs by misguided attempts to solve it. These designs cause risks, inefficiencies and regulatory responses that are far more costly than any likely mistake in the provision of adequacy. Of course, the adequacy problem needs a solution, but it should not exacerbate already difficult aspects of electricity markets. The misconception most responsible for the current state of affairs is the notion that a cleverly designed "energy-only" market can induce optimal adequacy, or something close to it, even while the market has insufficient demand elasticity. Interestingly, when the notion of reliability markets was first developing back in the late 1990s, the importance of adequate demand response in an "energy-only" market was fully recognized.

> In this approach there is no price cap that limits the market price. It is assumed that the elasticity of demand to prices is enough to prevent any occurrences of market's failure to supply because of lack of generation adequacy. —Pérez Arriaga (1999)

Unfortunately, when the idea of reliability options spread to the US, these assumptions were sometimes replaced with the idea that options priced only on the basis of their financial cost would solve the reliability problem in spite of a lack of sufficient demand elasticity.

In an ideal market, with sufficient demand elasticity, the market always clears. This means there can be no adequacy problem because involuntary load shedding occurs only when the market fails to clear and demand exceeds supply. In a market that always clears, energy prices do not and cannot determine the level of reliability for, with respect to adequacy, the market is automatically 100% reliable. Instead, energy prices do what every economics text says they do, they determine the efficient (not reliable) level of capacity. Any less capacity would cause expensive voluntary load reductions, and any more capacity would mean that too much had been spent on capacity. The crucial point is that no energy-only market, even with ideal demand elasticity, can solve the adequacy problem. No energy market, on its own, can ever answer the question: "What level of capacity provides optimal reliability?"

Of course if reliability is sold as a product, so that the customer can pay more and get more reliability, the market can answer the reliability question. But that is not an energy-only market, and it requires technology that is not yet in place. At present, you cannot buy more reliability than your neighbor, because you are on the same physical circuit and will always be blacked out together. Those who believe an energy market can solve the adequacy problem have simply misunderstood the economic theory of optimal investment. In a competitive market, optimal investment has nothing to do with reliability.

Energy markets fall into one of two categories: (Case 1) they always clear and have no adequacy problem, or (Case 2) they can fail to clear and do have an adequacy problem. To classify them correctly it is necessary to consider how each market would perform without market power or regulation. Case 1 occurs when there is enough demand elasticity. With ample demand response, there will be a long-run capacity equilibrium, but it may not be efficient if the price sometimes exceeds the Value of Lost Load (VoLL). Then the capacity level will be inefficiently high. Case 1 also includes the likely possibility that spot prices will approach, but not exceed VoLL. Then, both risk and market power are almost certain to reach problematic levels. If so, the regulated, reliability market described here is still likely to outperform an unregulated energy-only market.

Case 2, in which reliability is not guaranteed to be 100%, is the case considered in this paper. In most markets, few observers are willing to guarantee that involuntary load shedding is out of the question if the market is left to its own devices. Even so, some markets appear to be reliable on there own. Does this mean competitive energy prices are coming close to solving the reliability problem? Not at all. There are three possible explanations. First, suppliers may be exercising enough market power to attract new entrants. Monopoly power can easily produce more than enough reliability, but the resulting level has nothing to do with efficiency. Second, reliability can be the result of unrecognized regulatory interventions, such as tampering with the demand for operating reserves. Third, it is possible, though unlikely, that LSEs are expressing their own demand for reliability in a market that, in effect, is selling reliability to LSEs. For such an LSE-reliability market to work, the system operator would need to monitor LSE purchases of power, and the real-time performance of their suppliers, and cut off LSEs that have purchased insufficient or defaulting supply. We do not believe system operators are currently basing load shedding on contract positions, real-time purchases by LSEs, and the real-time performance of their suppliers. Even if this were the case, reliability would not be determined by consumers' value of lost load, but by the LSE's value of lost load, which would reflect the likelihood and severity of regulatory punishment—not a market signal from consumers.

To summarize Case 2, the case in which there is an adequacy problem, the reliability level is determined by regulators or market power and not by competitive energy prices. In fact, industry consultants have argued for years that electricity markets need market power to allow suppliers to cover fixed costs. This approach is common. But so are regulatory adjustments to guide the reliability level. When dispatchers (the engineering arm of the regulator) are worried about adequacy, they find more reasons to allow prices to rise and are more generous in their demand for operating reserves. When they see the system as overly reliable and prices as too high, they find ways to suppress the spot price. If the reliability problem grows more severe, regulatory intervention becomes more open, sometimes resulting in the purchase of capacity. When market power becomes more severe, regulatory intervention is also likely but it works to lower prices and decrease reliability. This is natural and indeed beneficial, but it is not exactly what economics means by a competitive market.

The common "proof" that a "pure-market" design is doing a "pretty good job" of letting the market determine reliability goes as follows. The lights are staying on, and there is no capacity market, therefore "the market" must be choosing "about the right level" of capacity. The flaw in the proof is that it overlooks both market power and regulators. In fact, "pure-market" approaches, are simply inadvertent hoaxes. But the hope for an energy-only, pure-market solution, is more deeply flawed than this mechanistic analysis indicates. The concept of an energy-only market solving the reliability problem without selling a reliability product, is logically impossible. It suggests that "the market" can do something "fairly well" when logic shows it cannot do it at all. Here is the argument that "energy-only" competitively determined reliability is simply impossible.

If nothing is known of consumers' utility for reliability, no limits can be put on the level of optimal reliability. Current markets have no access to information concerning how consumers value reliability. In other words, consumers take no market actions that are even partially based on reliability considerations. This is obviously true for consumers that do not have real time meters and who cannot be individually interrupted. These customer would be foolish to cut back when prices are high, because they cannot receive credit for their efforts, and they would be foolish to pay more for reliable service since they cannot physically be given more reliability than their neighbor who does not pay for more reliability. Many large consumers do have real-time meters and some can be interrupted. But, system operators are not prepared to black them out based on the performance of their contractual arrangements for power and the performance of their suppliers. If they are never blacked out for paying too little, they will not pay for reliability. So the market receives no signal of how they value reliability. They may well reduce their demand if the spot price rises to say $1000/MWh, but this tells nothing about the costs they would suffer if they were blacked out. Their price response is a step toward Case 1, in which reliability is prefect, but it provides no information about their VoLL.

In summary, customers currently reveal nothing about the value they place on reliability. Without this information, it makes no sense to think that a market can approximately determine how much reliability they would be willing to pay for. Currently, reliability is determined entirely by market power and regulators. Neither is ideal but at least regulators may attempt to solve the problem. But even unadorned market power is better than one particular "pure-market" design that has become popular in Australia and in the Western US. That design calls for increasing market risk for consumers so that generators can make more money selling them hedges. Although this can increase generator profits by assisting their exercise of market power

in forward markets, it has nothing to do with determining optimal reliability, and is certainly not suggested by any economic theory. This approach is an attempt to regulate the market power in forward markets to a higher level as a way of producing the right level of investment and reliability. One cannot get much further from the idea of competitive market efficiency.

2 The basic reliability-options solution

As just explained, if a market has an inherent reliability problem, then it must be solved by regulation or market power. The goal then is to design a regulatory approach that mimics a competitive market as closely as possible. By setting the spot market price cap at VoLL and using options to suppress risk and market power while preserving marginal incentives, regulation can be confined to the single task for which it is needed: determining the adequate level of capacity. Of course, reducing market power moves the market closer to the competitive ideal, but what is often overlooked is that reducing risk does too. While, present energy markets are risky because of price spikes, these are caused by the extremely low demand elasticity. If instead, the market did not have demand-side flaws, demand elasticity would be far higher, and risk far lower. Reducing risk moves the market towards this competitive ideal.

We will now describe a simplified design that illustrates the principles involved. Related approaches are described in Bidwell (2005), Chao and Wilson (2004), Cramton and Stoft (2006, 2007), Oren (2005), and Vazquez et al. (2002). It is assumed to be implemented in a single isolated market. Later we will discuss the problem of trade. At first, the energy market will be assumed to contain only the real-time market, but later, settlements that include forward markets will be described.

Step one of the design is to set the price cap on the spot market to the best estimate of VoLL. Of course this is a poor estimate, but any value in the 3,000 to €30,000 range should provide good dispatch incentives. Since VoLL will play no role in determining investment, the level is not critical. The difference in energy supplied under a €3,000 cap and a €30,000 cap is minuscule, because, with market power well controlled, these prices rarely are reached and by the time the price reaches €3,000, there are not many generators that can provide power and are choosing not too.

Step two of the design is to provide load with a complete hedge in the form of a reliability option with a strike price of say €300. Load purchases the target quantity of physical capacity together with a load-following call option at the strike price. In this way, load is 100% hedged from energy prices in excess of €300. If the energy price in an hour is €1000, each supplier has a financial obligation to serve its share of load. Since deviations are priced at the €1000 spot price, the supplier continues to be motivated by the spot price, although both load and suppliers are hedged from price volatility above the €300 price.

The key point is that all generation still faces the spot price even though it is hedged. Suppose a supplier owns 100 MW of capacity. If it provides 80 MW of power for the hour in question and has a 90 MW obligation, it is paid €80,000 because the spot price is €1,000, but it must pay (90 MW) × €(1000 – 300) as a hedge payment. If it provides 90 MW of power, it is paid €90,000 and is obliged to make the same €63,000 option payment. If it produces 100 MW it is paid €100,000, and again makes the same hedge payment. For every MW it increases or decreases its production, its net revenue increases or decreases by €1,000.

4

Note that when the spot price is €300 or above, it makes sense for virtually every generator to be producing, since marginal cost typically is less than €300, and as long as the suppliers produce their share of load, they will earn the strike price for all of their output. In other words, a generator with average performance is fully hedged against spot prices above €300 by its physical generator. This approach greatly reduces the risk from weather related price fluctuations. It does not however reduce performance risk. Although all risk is costly and hence undesirable, performance risk cannot be eliminated without eliminating the performance incentive. Generators will argue for a long list of exceptions due to "acts of God," but virtually all of these are to some extent under the generators' control, and since performance risk is quite small, it is best to simply ignore these arguments. In the full price market of step 1, no one would dare argue for exceptions. No matter what the reason a generator is out of service when the price is €10,000/MWh, no one would consider paying it for power it did not produce due to an "act of God." This obvious fact has proven extremely difficult for suppliers to remember during market design negotiations.

To summarize our progress with the design, the first step assured that the market conforms to a classic competitive design, although it still includes the currently unavoidable demand-side flaws (which lead to the absence of a robust demand response), and the compensating intervention of a VoLL price cap. As is well known, with this cap, and ignoring risk and market power, the market will provide both optimal dispatch incentives and optimal investment incentives. Of course the regulator is in control of investment incentives, and can induce any level of reliability. But if the regulator sets the cap to VoLL (which cannot be determined by the market), investment will be optimal.

Step two preserves the dispatch incentives perfectly on the generation side, but destroys the investment incentive, because although generators face the spot price on the margin, their revenues are limited exactly as if there were a price cap at €300/MWh. Why is this progress? Two advantages are obvious: a dramatic reduction in risk and in market power. The third advantage is the elimination of the need for a regulatory determination of VoLL. This is an advantage because it opens the door to a better method of controlling the level of reliability, as described in step three.

Step three sets the payments to generators for providing reliability options just high enough to induce optimal investment and adequate capacity. It does this by holding annual auctions to purchase new capacity up to the level required for reliability. These auctions determine the price of reliability options that is just sufficient to induce the required entry. For example, with the strike price of €300/MWh this might result in the average annual loss of €40,000 of revenue per MW of capacity relative to the spot market of step one. This is often termed the "missing money." In this case, new entrants will bid the price of reliability options down to €40,000/MW-year. If the cost of constructing new capacity increases or decreases, due to environmental restrictions or new technology, new entrants will bid just enough higher or lower to maintain a normal rate of return.

The result is that the regulator controls the level of capacity, but the market controls the price of capacity and the type and quality of capacity built. Hence the regulatory intervention has been strictly limited to the determination of the one factor about which the market has little information—the adequate level of capacity.

Although the auction design requires care to address the potential exercise of market power, the following simple procedure would work quite well. Each September an auction is held for reliability options, ROs, which take effect on January 1, just over three years in the future. Existing generators may choose either to enter the auction with a zero bid, or not to sell ROs. New projects are allowed to bid without restriction. The regulator bids a demand curve that intersects the target adequacy level at the most recent RO price and slopes down to the right by 5% in price for each 1% increase in capacity. It slopes up from the same point by 20% for each 1% decrease in capacity. The auction is held using a descending clock procedure (see Ausubel and Cramton 2004). All accepted bids are paid the clearing price, but existing generation receives one-year contracts while new generation may choose any contract length up to seven years. Once a new generator's initial contract expires it becomes an "existing" generator. If no new generation is purchased in a given year, all existing generators that bid, have their contracts extended for one year.

There is one further rule, which assures full hedging and limits market power in both the spot and capacity markets. Any generator that does not sell ROs for its full capacity receives only the spot price capped at the strike price. In other words, such generators in effect provide the hedge without compensation. Some designs omit this rule. Objections to it are based on ideological grounds. We find these unpersuasive and prefer to rely on economic analysis. As explained by Hogan and Harvey (2000) during the California crisis, suppliers will not give up their market power for free by entering into long-term contracts. They realize that their market power is valuable, and will extract approximately the value of this market power before relinquishing it by selling ROs. This rule prevents the exercise of market power in the capacity market, as well as the spot market. Only permanent retirements can reduce the quantity of existing supply.

3   How reliability options work in practice

There have been concerns that the mandatory RO system requires a centralized day-ahead market such as found in the Northeastern ISO's of the US (PJM, ISO-NE, and NYISO). Fortunately these concerns are unfounded. There is also concern that the RO mechanism is incompatible with a high level of bilateral contracting. In fact there always has been a high level of bilateral contracting in New England and other markets where this design has been adopted. How ROs integrate with bilateral markets and trade with other markets is explained below.

Assume a 100 MW generator has sold a reliability option with a strike price of €300, and the Transmission System Operator, is handling the reliability market. There are also long-term bilateral markets, a day-ahead market run by APX, and a balancing market run by the TSO.

Currently, European markets can be thought of as two settlement systems. A generator sells total energy of $Q_{Forward}$ in forward markets, either bilateral or centralized like APX, and $Q_{Forward}$ is scheduled with the TSO. The generator then delivers an amount Q_{RT} in real time. The forward quantity is settled at the privately determined prices regardless of what is delivered in real time, while the deviation from the forward sale, $Q_{RT} - Q_{Forward}$, is paid the balancing market price, $P_{Balance}$. Of course the deviation can be negative, in which case the generator pays the TSO. So without reliability options or when the balancing price is below the strike price, the settlement works as follows:

$$\text{Generator Revenue} = P_{Forward} \times Q_{Forward} + P_{Balance} \times (Q_{RT} - Q_{Forward})$$

With reliability options, and the balancing price above the strike price, the RO is settled second in what is essentially a three-settlement system. Each supplier is responsible for a share of the real-time load that is proportional to the quantity of reliability options it has sold. For example if it has sold 100 MW of options out of a total of 10,000 MW of reliability options, it is responsible for 1% of the load in every hour. Call the supplier's share Q_{Share}. The three-settlement system works as follows:

$$\text{Generator Revenue} = P_{Forward} \times Q_{Forward} + P_{Strike} \times (Q_{Share} - Q_{Forward}) + P_{Balance} \times (Q_{RT} - Q_{Share})$$

If the generator supplies exactly its load share, so that $Q_{RT} = Q_{Share}$, then it is fully hedged against the balancing price. However if it deviates either up or down from its share, it is paid or must pay the balancing price. Hence its incentive to perform has not changed, and the balancing market continues to play its traditional role. As can be seen, there is not much interaction between the forward transactions and the reliability option. In particular there is no reason generators cannot sell all of their power in the forward markets and sell reliability options for all of their capacity in the capacity market. Three things have changed as a result of the hedge built into the reliability option: the average generator earns at most the strike price in the balancing market, both load and generation are less at risk, and the forward contract needs only to cover prices below the strike price.

Notice that the sum over all generators of Q_{RT}, the total power delivered, equals the total load, which equals the sum over all generators of Q_{Share}. This means the sum over all generators of $P_{Balance} \times (Q_{RT} - Q_{Share})$ is zero. These terms are the incentive payments for under and over performance by generators, relative to Q_{Share}. So these payments do not affect load, but are payments from poorly performing generators to the better performers. As a consequence of this fact, generators do not have to be concerned that the TSO will under-procure capacity because, on average, this cannot change the total performance payments to generators which always sum to zero. Under-procurement of capacity will only increase the number of hours when there is a shortage and generators are paid the strike price.

Another concern is that European markets are open to trade between countries. If the price is high in Germany, but Dutch generators are under reliability options, they might leave the Dutch market where, on average, they can earn only the strike price, and sell as much as they can into the German market. To examine this possibility, the settlement must include a term for exports. Since the concern is with the effect of reliability options, the real-time price must be above the strike price, so it is safe to assume a generator will produce the most it is capable of, Q_{Max}. Next note that only the power delivered to the domestic balancing market, $Q_{Domestic}$, receives the balancing price. The export quantity is then, $Q_{Max} - Q_{Domestic}$, and the settlement works as follows.

$$\text{Generator Revenue} = P_{Forward} \times Q_{Forward} + P_{Strike} \times (Q_{Share} - Q_{Forward})$$
$$+ P_{Balance} \times (Q_{Domestic} - Q_{Share}) + P_{Export} \times (Q_{Max} - Q_{Domestic})$$

Because the derivative of Generator Revenue with respect to $Q_{Domestic}$ is $(P_{Balance} - P_{Export})$, the incentive to export is exactly the same as without reliability options. Hence there is no justifiable concern with a disruption of the balance of electricity trade, or a collapse of the domestic market.

Another practical concern is the assignment to load serving entities (LSEs) of responsibility for the cost of ROs. Because the options are procured by the TSO, the LSEs are not burdened

with purchasing ROs and need not make any long-term commitments by purchasing them. This is a great advantage because it means LSEs are at little risk from consumers moving from one LSE to another. The cost assignment is simply adjusted each year and based on the LSE's coincident peak load during the year. This determination is best made after the fact. For example, option cost responsibility from 2010 should be based on the peak loads during 2010. To reduce randomness in loads on any give day, a weighted average of the three highest peak-load days could be used.

Another practical concern is that load is not exposed to the full spot price because of the hedge. Although most load is still not on real-time meters, these are becoming more prevalent. Moreover, LSEs can implement various programs to encourage conservation during times of peak load, and it would be worthwhile to properly motivate them to do so. This can be accomplished as follows. First compute the peak energy costs, PEC, of each LSE. This is simply the integral of its load times $\text{Max}(0, (P_{Balance} - P_{Strike}))$ over the year. Then, since each LSE is assigned a reliability share, L_{Share}, based on its coincident peak load, this share can be used to compute its share of the total of all PECs. Each LSE then pays a penalty equal to the amount by which its actual PEC exceeds its share of total PECs.

$$\text{Peak-Load Penalty} = \text{PEC} - L_{Share} \times (\text{Sum of all PECs})$$

The sum of the penalties is zero, and the derivative of a load's penalty with respect to its own peak energy cost is $(1 - L_{Share})$, which is near one as long as LSEs are small. This means that purchasing a MW of power, when the balancing price is above the strike price, costs each LSE and amount $P_{Strike} + (1 - L_{Share}) \times (P_{Balance} - P_{Strike})$, which is very nearly $P_{Balance}$. In other words the penalty makes the LSEs face the balancing price on the margin even though they pay no more on average. There will be some risk to loads from this performance penalty but it is small and it is only what is inevitable if loads are to face the real-time price on the margin. They are still completely hedged against price spikes caused by weather, nuclear outages, or other events out of their control. This same technique can be used by LSEs to pass real-time price signals through to their loads equipped with real-time meters.

One final concern is that reliability options may impose burdensome new information requirements. However, the TSOs are already aware of the quantities transacted in the forward markets because these must be scheduled. They are also aware of actual production and of which generators are exporting how much power. Because the TSO will conduct the capacity auctions, it will know who owns the reliability options. It also knows the daily loads of the LSEs, and the balancing market prices. This is all the information needed to implement this reliability option design. In particular there is no need to collect more information about bilateral transactions.

4   Reliability options in a hydro-dominated system

If supply is mainly from hydro-electric generation, the limiting factor is not likely to be capacity (the ability to provide energy in peak hours) but rather, firm energy (the ability to provide energy in dry periods). As a consequence the TSO will need to purchase firm energy options. Just as capacity is the physical basis for reliability options, so firm energy options have a physical basis that involves a longer-term supply of energy. Firm energy is defined as the amount of energy a generator can deliver per month during an exceptionally dry period (a worst-case benchmark). A typical thermal unit is certified at its nameplate capacity times its average availability, such as 92% of nameplate. In contrast, a hydro resource's firm energy may be well

below its nameplate, say 35%, due to a limited water reservoir and low inflows during dry periods. A new unit's firm-energy contribution is how much less energy the system would have without the unit in the worst-case benchmark.

The second difference in a firm-energy market is the way load-share is defined when the real-time price rises above the strike price and the option comes into play. The total quantity of the call option follows load, and is divided between thermal and hydro resources as follows. The load obligation is first divided among the thermal resources in proportion to their firm energy ratings. Once load exceeds the total firm energy ratings of thermal capacity, the excess load is divided among the hydro generators in proportion to their firm energy ratings. Economically, this division of load between thermal and hydro generators makes sense, since it is consistent with efficient dispatch of the units. During scarcity hours the hydro opportunity cost sets the price and thermal resources run at capacity if available. One might think that this approach is biased against hydro resources, since hydro is asked to do more load following than the thermal resources. Indeed, there would be a bias if the ability to follow load was scarce; however, since there is a surplus of capacity, load following is not costly and the capability is efficiently priced at zero.

Given these definitions, settlement proceeds exactly as it does in the reliability-options market just described.

In a market in which it is unclear whether capacity or firm energy is the scarce resource to achieve generation adequacy, the approach can be enhanced to accommodate both products. Each resource is rated for both its capacity and its firm energy, and offers the two products as a package. There is a target for each product. The descending clock auction then has two prices, one for capacity and one for firm energy. The prices descended until a supply and demand balance has been achieved in both products. In a hydro-dominated system, this will imply a zero price for capacity and a positive price for firm energy. In a thermal-dominated system, it will imply a positive price for capacity and a zero price for firm energy.

5   Implementation in New England and Colombia

Reliability markets based on this reliability options approach have been adopted in New England's thermal-dominated market (Cramton 2006) and Colombia's hydro-dominated market (Cramton and Stoft 2007). The markets are currently in a transition period in which the capacity and firm energy prices are set administratively. Both markets will have their first auctions in early 2008. Both markets differ somewhat from the specific approach described above.

New England has locational pricing, so it was important to implement the capacity market on a zonal basis, so that adequacy would be achieved throughout New England. Another important difference is the way that market power is addressed in the capacity market is more complex than what we have suggested above.

Colombia's firm energy market is quite similar to what we describe above. The strike price of the reliability option is only about €100/MWh. This is still above the marginal cost of nearly all generation, which means that suppliers have a physical hedge to protect against prices above the strike price. Another change is a longer planning period of four years, rather than three to accommodate longer-lead time hydro projects. Indeed, to accommodate large hydro products,

suppliers can sell firm energy up to seven years ahead at the four-year-ahead price, subject to some restrictions.

6   Risk analysis based on market data

In both New England and Colombia, we have examined the supplier risk associated with the reliability options approach. We have examined company risk using historical market data, as well as detailed market simulations carefully calibrated to the specific market setting.

The findings in Colombia (Cramton et al. 2007), where we did more detailed simulations, are as follows.

- Lumpy investment means that few new units are added each year. Indeed, in 27% of the years no new entry occurs. (This is an overestimate to the extent that the size of proposed projects reflect the actual need in the year, as one might expect.)

- The mandatory hedge is remarkably successful in reducing risk. In the benchmark case, the hedge reduces aggregate profit risk by a factor of 7. More importantly, the hedge reduces company risk by a factor of 4.5 in the benchmark case. Even when we assume a high level of demand response so that prices remain low during scarcity periods and there is less profit risk to start with, the hedge reduces company risk by 55%.

- A higher strike price increases risk. Increasing the strike price shifts the profit distribution toward the no hedge case (a strike price of infinity). This results in a large increase in energy rent risk and a small decrease in hedge payment risk. The overall impact is a large increase in profit risk.

Taken together, the simulation results demonstrate the risk reducing benefits of the reliability market. Provided there is competitive new entry in response to load growth, the reliability market should work well at coordinating investment in new supply, while minimizing supplier and consumer risks.

7   Conclusion

The forward reliability market approach described here is the product of a systematic development based on clear economic principles. Indeed, once it is understood it seems almost inevitable. The first step is the classic VoLL pricing system which is known to be both short and long-run optimal if VoLL is known and risk and market power are assumed to be costlessly suppressed by unspecified means. The second step is to suppress risk and market power, the two evils of VoLL pricing, by introducing reliability options. These do not interfere with real-time price signals, so on the second step, the market retains the dispatch optimality of classic VoLL pricing, but without the problems.

The second step suppresses a significant amount of generation revenue and thereby destroys the investment incentive, so the third step restores it by introducing a capacity market, which induces an adequate level of investment by procuring the appropriate quantity of reliability options. All generation, new and existing will want to sell reliability options for their full capacity because these options fetch a high price relative to the financial cost of the option. Participation in this market is guaranteed by the rule that non-participating generators receive the spot price capped at the option strike price.

The auction for procuring reliability options takes place three years in advance of the effective date, so that there is time for new entry to back the options. To suppress market power in the reliability-option auction, only new capacity bids are allowed to set the price. The auction is a descending-clock auction. By inducing investment with an auction instead of high prices, not only does the regulator have better control of the average reliability level, but far better investment coordination is assured.

Investment coordination prevents the boom-bust cycles which increase both investor risk and reliability risk for consumers. Without an auction, as the market tightens it offers an increasingly large prize for the next entrant. However, entry is a secretive process, and so simultaneous entry is possible. Aware of this, investors are torn between holding off until the prize is large enough to support some simultaneous entry and entering quickly to ward off competition. The optimal strategy is mixed and the outcome chaotic. With an auction, the TSO coordinates entry without reducing competition. There are multiple simultaneous bids, but the TSO selects only as many as needed. This stability benefits both consumers and investors.

Although there are many concerns about the use of reliability options, with the designs specified here, none prove warranted. There is no difficulty deciding which private contracts are acceptable substitutes for reliability options, because no substitutes are accepted. One hundred percent coverage by reliability options does not interfere with 100% coverage with bilateral contracts. Reliability options provide price coverage above the strike price; bilateral contracts provide price coverage below the strike price. Although reliability options limit the average real-time price to the strike price, the marginal price for both load and investors remains the spot market price. This preserves incentives and prevents any increase in exports relative to the present system even on days when other countries have high prices. The benefits of this design are significant. The design minimizes risk and market power, while coordinating efficient entry.

References

Arriaga, Ignacio J. Pérez (1999), "Reliability in the New Market Structure," IEEE PES 1999 Summer Meeting, Plenary Session.

Ausubel, Lawrence M. and Peter Cramton (2004), "Auctioning Many Divisible Goods," *Journal of the European Economic Association*, 2, 480-493, April-May.

Bidwell, Miles (2005), "Reliability Options," *Electricity Journal*, June.

Chao, Hung-po and Robert Wison (2004), "Resource Adequacy and Market Power Mitigation via Option Contracts," EPRI, Palo Alto, CA.

Cramton, Peter (2006), "New England's Forward Capacity Auction," Working Paper, University of Maryland.

Cramton, Peter and Steven Stoft (2006), "The Convergence of Market Designs for Adequate Generating Capacity," white paper for the California Electricity Oversight Board, March 2006.

Cramton, Peter and Steven Stoft (2007), "Colombia Firm Energy Market," *Proceedings of the Hawaii International Conference on System Sciences*.

Cramton, Peter, Steven Stoft, and Jeffrey West (2006), "Simulation of the Colombian Firm Energy Market," Working Paper, University of Maryland.

Hogan, William W. and Scott M. Harvey (2000), "California Electricity Prices and Forward Market Hedging," Working Paper, Harvard University.

Oren, Shmuel S. (2005), "Generation Adequacy via Call Option Obligations: Safe Passage to the Promised Land," *Electricity Journal*, November.

Vazquez, Carlos, Michel River, and Ignacio Perez Arriaga (2002), "A Market Approach to Long-Term Security of Supply," *IEEE Transactions on Power Systems,* 17 (2): 349-357.